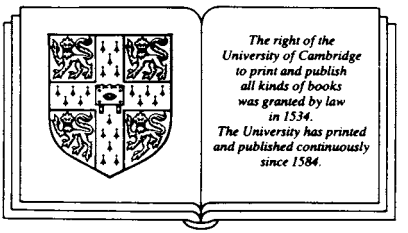
The Structure of Emotions

INVESTIGATIONS IN COGNITIVE PHILOSOPHY

Robert M. Gordon



Cambridge University Press

Cambridge
New York New Rochelle Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1987
First paperback edition 1990

Library of Congress Cataloging-in-Publication Data
Gordon, Robert M. (Robert Morris)
The structure of emotions.
(Cambridge studies in philosophy)
Bibliography: p.
Includes index.
1. Emotions. I. Title. II. Series.
B815.G67 1987 152.4 86–28397

British Library Cataloguing in Publication Data
Gordon, Robert M.
The structure of emotions: investigations in cognitive philosophy. – (Cambridge studies in philosophy)
1. Emotions
I. Title
152.4 BF531

ISBN 0 521 33164 1 hardback
ISBN 0 521 39568 2 paperback

Transferred to digital printing 2003

Contents

1

Formal insight

On July 1, 2020, the centuries-old debate was settled by decree: *There are no emotions.* Thenceforward no one was, nor ever had been, in a state of "anger." The same fate befell fear, joy, embarrassment, amusement, grief, and all the rest: Banished from discourse, in private thought policed by (dare we say it?) guilt, they became unmentionable and unthinkable. For all the furrowing of brows, the narrowing of eyes, the clenching of fists (sometimes precursors to nastier actions), anger and its kin were, in the words of the decree, "Paleolithic fictions." The general terms 'emotion' and 'emotional' were, of course, outlawed as well – though these were hardly to be missed, save by a few philosophers and here and there an unenlightened psychologist.

Objectivity ruled.

HEBB'S EXPERIMENT

One hardly needs to speculate on the consequences of this fanciful ban. Something like it has already been tried, though in a less critical context: in the scientific description of chimpanzee behavior. A ban on talk of anger, fear, and joy in the higher ("anthropoid") apes may, of course, be better motivated than the one imagined above. Some of the publicity given to recent attempts to teach language to chimpanzees and gorillas suggests that many people, including even some scientific investigators, have a thirst for seeing humanlike attitudes and emotions, a perceptual bias verging on primitive animism. Why not, then, adopt at least experimentally the rule *Thou shalt not anthropomorphize the anthropoids*?

The psychologist D. O. Hebb reported the results of just such an experiment at the Yerkes Laboratories of Primate Biology (1946). His account, published more than forty years ago, remains

fascinating, and I find Hebb's discussion – his analysis of the human aptitude for "recognizing" emotions in the behavior of people and animals – philosophically more sophisticated than the bulk of philosophical writing on the emotions in the intervening decades.

Daily records were kept of the behavior of a number of chimpanzees, with the aim of pointing up significant behavioral differences between one individual and another – differences that might guide new staff members in their dealings with the animals. Over a period of two years, however, the records were to be kept devoid of all attempts to "anthropomorphize." Observers were enjoined to describe the behavior of the apes without recourse to the mentalistic vocabulary of what is sometimes called "commonsense psychology" (or, typically for its less favorable connotation, "folk psychology"). One might report that the animal moved its limbs in certain ways, perhaps even that it "attacked" an attendant; but there was to be no mention of any alleged thoughts, beliefs, or desires. Most important, there were to be no attributions of emotions and emotional tendencies: The animal must not be said to have attacked out of "anger" or "hostility."

According to Hebb, the records kept over those two years were far less helpful as a guide to new staff members than records drawn from the earlier "anthropomorphizing" years. It was impossible to learn which apes may safely be approached, by whom, in what manner. "All that resulted was an almost endless series of specific acts in which no order or meaning could be found" (1946:88).

During the anthropomorphizing years, on the other hand, distinctions useful for the prediction of behavior could be readily made:

> By the use of frankly anthropomorphic concepts of emotion and attitude one could quickly and easily describe the peculiarities of the individual animals, and with this information a newcomer to the staff could handle the animals as he could not safely otherwise. (1946:88)

One example Hebb cites concerns two chimpanzees, Bimba and Pati, who on many occasions behaved in ways that were "superficially" quite similar. Each made violent assaults on their keepers, behavior that observers were unable to distinguish by purely "behavioristic" description. Analysis of the "objective" records revealed no significant differences in the circumstances, frequency, manner, or severity of their attacks.

Nevertheless, those who had known the two animals over a long

period, particularly those who had observed their behavior at times when they were *not* attacking, were unanimous in attributing all of Bimba's attacks to *anger* and Pati's to a *general malice* or *hatred of man*. The basis for this distinction, Hebb surmised, was that at other times Bimba "is always responsive to man, and acts in a way which promotes contact and petting by the attendants," whereas Pati is at best unfriendly. Some observers had indeed *thought* they could perceive a difference in the nature of the attacks by the two animals, but this was apparently a "halo effect," an attribution wholly due to the differences they had seen at *other* times:

> Bimba's attacks seemed to occur only after some movement by the attendant that might appear like a threat, or like teasing; and it seemed that her attacks were always open, with the frank violence of anger, while Pati's were stealthy. But this was a fallacy of memory. (1946:92)

Although Hebb did not spell out the practical consequences of this distinction between Bimba's and Pati's attacks, one can imagine what some of these might have been. A prudent attendant, bearing in mind that Bimba, though friendly to man, is easily provoked to anger, would try to avoid slighting her or giving her insufficient attention – two of the provocations the attendants had mentioned. Pati, on the other hand, one had best avoid altogether, unless equipped with protective padding and a plan for escape. For it is in the nature of anger that it arises from – is "provoked by" – certain specific types of situations (or "cognitions"), such as a "slight," whereas hatred is a long-term disposition that, once established, needs no provocation at all.[1] (Still other "problem" chimpanzees – those that were shy, or afraid of strangers – would be best approached in a way that permits gradual "desensitization.")

Hebb shows quite effectively that superficially similar behavior sequences may be symptomatic of different underlying states, and consequently may presage different behavioral possibilities for the future. His experiment suggests that anthropomorphic description,

1 See Aristotle's definition of anger: "an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification towards what concerns oneself or towards what concerns one's friends (felt toward some particular individual, not man in general)" (1924: Book II, chap. 2). But "Whereas anger is always concerned with individuals, hatred may be directed against classes, e.g., any thief, any informer. Anger may be cured by time, hatred cannot. The angry man wants his victims to feel, the hater does not mind whether they feel or not; anger is accompanied by pain, hatred is not" (1924: Book II, chap. 4).

particularly the attribution of emotions and character traits, can be useful in allowing us to see, or at least to surmise, the divergent states that may lie "beneath" outwardly similar behavior.

Of course, one would be ill-advised to enter an ape cage confident that everything one believes about human emotions carries over to chimpanzees. Quickly one would learn to respect the toothy grin of an angry chimpanzee as betokening something other than amusement. No doubt, too, some scientific purposes would be better served by behavioral descriptions that avoided any implication of humanlike motives, emotions, or character traits. Yet Hebb concluded after his two year experiment that such "objective" descriptions had

> . . . missed something in the behavior of the chimpanzee that the ill-defined categories of emotion and the like did not – some order or relationship between isolated acts that is essential to the comprehension of behavior. (1946:88)

Our traditional taxonomy of emotions, he writes, "evidently implies an elaborate theory," and its practical value in predicting behavior suggests that there is "some truth in it," as regards chimpanzees as well as human beings (1946:97–98). "Whatever the anthropomorphic terminology may seem to imply about conscious states in the chimpanzee, it provides *an intelligible and practical guide to behavior*" (1946:88).

BEYOND BELIEFS AND DESIRES

Hebb was not alone in thinking that commmonsense psychology is a valuable guide to behavior – suitable not only in everyday experience but also, with refinements, in the scientific study of behavior. At least in the study of *human* behavior, some of the concepts of commonsense psychology, along with part of its theoretical structure, appear to have been widely adopted by the behavioral sciences. Alvin Goldman has argued persuasively that

> . . . much of the work done in the behavioral sciences either presupposes concepts quite similar to those of wanting and believing or frames hypotheses which are compatible with the operation of wants and beliefs. (1970:131)

Among the scientific or technical terms Goldman cites as typically meaning something like 'want' or 'desire' are 'motive,' 'need,'

'goal,' 'attraction,' and 'utility.' Terms that mean roughly the same as 'belief' are 'cognition,' 'expectancy,' and 'subjective probability.' (Goldman adds the weaker disjunct, "frames hypotheses which are compatible," chiefly to accommodate behaviorists, such as B. F. Skinner, who profess to give explanations that are *in*compatible with belief-and-want explanations. Goldman argues against such incompatibility.)

The so-called emotions have obviously not fared as well as beliefs and desires (or wants). Commonsense terms such as 'fear,' 'anger,' 'delight,' 'pride,' and 'embarrassment,' do not have their scientific counterparts, not at least in wide use; and the "elaborate theory" Hebb finds in our traditional taxonomy of emotions is rarely represented in scientific theories of behavior. Of course, scientists who use *belief*-like concepts to explain behavior may be interested in commonsense *beliefs about* such putative states as fear, anger, delight, pride, and embarrassment. For such beliefs, even if they are thought to have no place in scientific theorizing, surely figure in commonsense theorizing about behavior (one's own as well as that of others); and such theorizing is, arguably, a particularly important *influence* on behavior (see Heider 1958).

In sum, behavioral scientists have generally tried to get by without adopting and invoking the concepts of the various emotions. If there is an elaborate commonsense theory of such states, few scientists have been willing to give it their professional endorsement. One reason, no doubt, is lack of agreement as to the content of that theory, and hence as to what they would be buying into. A second motive for withholding endorsement is parsimony: If one can make do with fewer mental states, then one ought to. Why introduce talk of anger, joy, embarrassment, and fear if any behavior explained by these states can be adequately explained as a product of beliefs and desires alone?

Is there any reason, then, to deny the explanatory adequacy of beliefs and desires alone: to think that an explanatory scheme that embraces beliefs and desires, or their technical counterparts, could not get by without incorporating the various emotions as well? I shall answer with a qualified "Yes." Two fictional examples will be discussed: one concerning an emotion typical of those discussed in Chapter 3 (the "factive" emotions), the other concerning two "epistemic" emotions discussed in Chapter 4.

The book burning. Jones, the department chairperson, is having an affair with Smith, one of the new assistant professors. Jones is an unmarried woman; Smith, a married man. She gives him her book manuscript for comment. Smith takes it home and leaves it on his desk. The two drive to a conference in another city. Smith's wife burns the book manuscript, one page at a time.

Why did she do that? What *desires* might figure in the explanation of her behavior? Trivially, this might be one:

(*Db*) a desire to burn Jones's manuscript.

One might, of course, add that she believed a particular object to be Jones's manuscript; that is why she burned that object. But this is not likely to put anyone's mind to rest. The wish to have Mrs. Smith's book burning "explained" would probably not be satisfied by a belief-and-desire explanation that failed to explain why she desired to burn Jones's manuscript.

Such an explanation would no doubt involve

(*Ba*) a belief that Jones and Mr. Smith had been having an affair.

But what is the connection between *Ba* and *Db*? Why would *Ba* make her want to burn Jones's manuscript? A relevant presumption would be that she believed the destruction of the manuscript to be a bad thing from the point of view of the chairperson (and/or the husband), a setback or at least a slight. We are likely to explain *Db* by such a belief together with

(*Ds*) a desire to do something that would be a setback or a slight to the chairperson (and/or the husband).

Thus we explain what Mrs. Smith did by *Db*, and we explain *Db* by *Ds*. But how do we explain *Ds*? What is the connection between *Ba* and *Ds*? One possibility is that she had a desire to deter the two from continuing the affair. She believed that a setback or a slight in response to the affair might lead them to expect even worse if the affair continued. This might explain her burning the manuscript, though perhaps not her doing so one page at a time.

But suppose we add a further complication. On the way back from the conference, Jones and Mr. Smith die in an automobile accident. After their funerals, the widow burns the manuscript. Here one can't explain *Ds* by a desire to deter the lovers from continuing the affair. We might explain it by attributing to Mrs.

Smith a desire *to get even*: perhaps, by "evening the scales," to restore an abstract retributive "justice," or perhaps to restore her "honor" and thereby her self-respect. Even a posthumous setback or slight might seem, to some minds, to set things right again.

But one may find her behavior plausible and even "understandable" without imputing such a principle of posthumous equilibration. She needn't have been acting on principle at all. *Ba* could lead to *Ds* by a quite different path. Presumably she also had

($D^{\sim}a$) a retrospective wish that her husband and the chairperson *had not* been having an affair.

But *Ba* and $D^{\sim}a$ jointly constitute what may loosely be called a "wish-frustration": a belief that something is the case together with a wish that it weren't. (This usage is explained in Chapter 2.) And human beings are so constructed, it appears, that wish-frustration *frequently has notable effects*. It may have observable physiological effects and effects on one's competence in various tasks; more important, it may cause one to have desires one would not otherwise have had, and thus to be motivated to act in ways one wouldn't otherwise have acted. A particularly noteworthy effect is that of motivating one to do something to harm the person or persons who caused the wish-frustration. Wish-frustration often has this effect independently of any "deontological" principles of justice, duty, or honor and any instrumental aims such as deterrence. This characteristic motivational pattern approximates what we call "anger" (albeit "minimal" anger, without the motivational and communicational complexities that often enter in). Thus one can explain *Ds* simply by supposing that the wife was still *angered* by the affair, perhaps even more so now that it seems to have been a factor in her husband's death.

To say that human beings are sometimes apt to be angered, embarrassed, or ashamed is to say little more than that among members of our species there are certain typical effects, particularly on motivation, apt to be produced by wish-frustration, with particular types of wish-frustration tending to cause particular types of effects. (Not all the so-called emotions are effects of wish-frustration, of course; but as it happens, the majority of those for which we have names are. These are among those discussed in Chapter 3, "Factive Emotions.") Unless these motivational effects are recognized, there will remain, as we have seen, serious gaps in

our attempts to explain human behavior in terms of beliefs and desires or their scientific counterparts.

The two farmers. Suppose that two farmers each wish that it would rain, so that the crops will not be ruined by drought. Each believes as strongly as the other that his crops will not survive another week without water, and each cares as much as the other about the survival of his crops. Farmer *A* sets out pipes in preparation for irrigating the land in case it doesn't rain. Farmer *B*, however, takes no such measures.

Such differences in action would be readily explainable if we could suppose that *A* believed it would not rain whereas *B* believed it would. Given the attitudes we are supposing both farmers to have, *A acts as if* he believed it would not rain: His dispositions to behave are, at least in most important respects, just like those of a person who was fairly sure it would not rain. And *B acts as if* he believed it *would* rain. But let us suppose that both are uncertain whether it will rain or not: They have heard from a source they trust that there is, let us say, a fifty percent chance of rain within the week. So neither believes it *will* rain; neither believes it *will not*. Nor do they differ in any other relevant beliefs.

The story is not at all implausible. Yet if motivational differences could be accounted for only by differences in beliefs or desires, the story would be *impossible*. But there is at least one other way to explain the difference in their behavior. For one can suppose that whereas farmer *A* is *afraid it will not* rain, farmer *B* is *hopeful it will*. One who fears or is afraid it will not rain will tend to act and feel like a person who *believes* it will not rain (but wishes it would). Likewise, one who is hopeful it will rain will tend to act and feel like a person who *believes* it will rain (and wishes it to).[2] There is an important motivational (and thus functional) similarity among fearing, being hopeful, and believing. Once again, serious explanatory gaps remain in any belief-and-desire psychology that fails to recognize these motivational analogues of belief. (These emotions are discussed in Chapter 4, "Epistemic Emotions.")

It must be admitted that, as is always the case, there are alternative

2 I distinguish *being hopeful* from *hoping*. One may hope without being hopeful; and it is the *hopeful* person who tends to act as if with belief. The two farmers example is discussed at greater length in Chapter 2.

ways of bridging explanatory gaps in belief-and-desire psychology. One might impute to the two farmers differing attitudes to *risk*, for example: *A* is more conservative, whereas *B* is more tolerant of risk. Thus *A* needn't be afraid it won't rain; he is simply cautious. But suppose that we have observed the farmers' behavior at other times, and seen no evidence of such a difference – or even seen evidence that it was *B* who tended to act conservatively. (Unlike in the case of Bimba and Pati, whose behavior "at other times" indicated that their attacks were differently motivated, in this case the evidence "at other times" tends to *rule out* certain differences in motivation.)

One might still avoid the fearful-hopeful distinction, for example, by speculating that farmer *A* had undergone a sudden conversion to conservatism or that *B* had undergone a complementary change in character. Or perhaps *A* has always been the more conservative *with respect to the survival of his crops in times of drought*; it's just that his attitude toward risk of this special sort had never before come to the test, this being the first drought in his experience. But in seeking alternatives to "emotional" explanations we are now being forced to "fetch" very far, straining credibility. (For an account of what it is for an explanation to be "farfetched," see Chapter 7.)

If one is to explain or predict human behavior in terms of beliefs and desires, then one should be prepared to introduce emotions as well into the explanatory scheme. That is the point I have tried to illustrate with the examples of the book burning and the two farmers. Of course, one can, as we have seen, find multitudinous ways around explanatory gaps in belief-and-desire psychology, if one is willing to fetch far enough. And that allows much slack to parsimonious theoreticians who would prefer to extract from folk psychology only the bare minimum of theoretical commitments. If parsimony is deemed more important than credibility, then one can get by without the emotions. This may partially explain why so few behavioral scientists have felt pressure to introduce the emotions into their theories of the springs of human behavior.

PHILOSOPHICAL INSIGHT

Remarkably many of the major classical philosophers took it as a major challenge to their analytical skills to attempt definitions of

the various emotions: perhaps most notably, Aristotle in the *Rhetoric*, Descartes in *The Passions of the Soul*, Hobbes in the *Leviathan*, Spinoza in his *Ethics*, and Hume in *A Treatise on Human Nature*. What they were doing in their defining, I suggest, was to make explicit the elaborate commonsense theory that Hebb thought he saw in "our traditional taxonomy of emotions."

This is, of course, a theory that is far more widely *used* than *articulated*. When these philosophers come close to success in articulating it, they touch a nerve: Their definitions strike readers as apt, insightful, and revelatory. Consider, for example, a few of Spinoza's definitions (1883):

> *Regret* is the desire or appetite to possess something, kept alive by the remembrance of the said thing, and at the same time constrained by the remembrance of other things which exclude the existence of it.
>
> *Consternation* is attributed to one whose desire of avoiding evil is checked by amazement at the evil which he fears.
>
> *Fear* is an inconstant pain arising from the idea of something past or future, whereof we to a certain extent doubt the issue.

These definitions impress us, I suggest, not because they nicely capture the nuances of ordinary expression – for they do not – but because they seem to tell us something about ourselves. Spinoza is explicit on this point:

> I am aware that these terms are employed in senses somewhat different from those usually assigned. But my purpose is to explain, not the meaning of words, but the nature of things. (1883:178)

The classical definitions may be seen as answers to questions of the traditional Socratic form: 'What is regret?' 'What is consternation?' and so forth. Taken in the abstract, such questions might be thought to concern the meanings of certain words. But philosophers rarely pose Socratic questions in the abstract. Socrates himself, it is clear, had hoped that his own techniques for answering such questions as 'What is justice?' and 'What is piety?' would provide guidance, not merely in the use of certain "buzz words" but in practical matters of the greatest importance. Thus he asked, "What is it to live a *just* life?" for example, on the understanding that, once we have discovered the correct answer, no rational doubts would remain as to *how to live* one's life.[3]

3 Socrates is called upon to defend this assumption in Book I of Plato's *Republic*.

So, too, in asking what regret or fear is, the philosophers did not pose their questions in the abstract. Rather, they took it for granted that, whatever the answers to these questions may be, *human beings are in fact susceptible* to fears, regrets, and the like. Given this assumption, to discover what fearing and regretting are is to discover something about the susceptibilities of human beings. It is also, *perhaps*, to discover something about chimpanzees and walruses and groundhogs, given the more questionable assumption that they too are susceptible to fears and regret; and even, perhaps, about nonmammalian species and emotional robots of the future. But if any beings are susceptible, surely we human beings are. Given the mere assumption, then, that some beings are subject to fear and regret, to discover what fear and regret are is to discover something about ourselves. (The claims I am making for philosophical analysis are certainly not novel. Those readers who need no convincing might wish to skip to the end of this section. For others I offer a further illustration, taken from a different area of philosophy.)

An example drawn from the heyday of "linguistic philosophy" shows quite plainly how a philosophical answer to a 'What is?' question can at least seem to be telling us something about ourselves. In a classic paper on "Meaning" (1957), H. P. Grice tried to explain (among other things) what it is for a speaker to "mean" something by an utterance. According to Grice, one means something by an utterance if and only if one intends it to have a certain effect on an audience and further intends that the effect be achieved in a certain way. More specifically, what is intended is that the effect occur as a result of the audience's *recognizing* the speaker's intention to achieve that effect.[4]

Expressed in somewhat different language, Grice's "analysis" can be seen as a hypothesis that certain *inferential processes* go on in the planning of speech behavior. Roughly translated, the thesis is that when we have meant something by a particular utterance *U*, the utterance was "selected" on the basis of two complex predictions:

4 Subsequent commentaries have shown that Grice's analysis must be modified in some important ways. There are more radical problems with Grice's further effort to show that a similar analysis could explain what it is for words or sentences to mean something (as opposed to a particular speaker meaning something by a word or sentence).

- That the utterance of *U* would cause the audience to infer that *U* was selected because of the speaker's prediction that the utterance of *U* would have a certain effect *e* on the audience
- That the utterance of *U* would indeed have effect *e* on the audience, precisely *because of* the audience's inferring that *U* was selected because of the speaker's prediction that the utterance of *U* would have a certain effect *e* on the audience

It is unlikely that many people, at least before having read Grice's arguments, would have a ready answer if asked whether their meaningful utterances were "selected" in this way. They could say with far greater assurance whether they had *meant* something by their utterance, and indeed just what they had meant. Thus the definiens in Grice's analysis is initially more problematic in its application than the expression it defines. The analysis, if we assume it to be correct, would tell us that on those more readily identifiable occasions on which the definiendum applies – for example, where someone means something by an utterance – the definiens also applies. Now, when we recognize that someone meant something by an utterance, we can be equally sure that the utterance resulted from inferential processes of the sort Grice describes. We can also be sure that corresponding inferential processes go on when one *understands* a speaker to mean something by a particular utterance.

Spinoza's definitions seem to offer insights into human nature, I suggest, in part because, like Grice's, they tell us *what is going on* on certain more or less readily identifiable occasions: when we experience or undergo what we call "regret," or "consternation," or "fear." The definiens in any of Spinoza's definitions of the various emotions is initially more problematic in its application than the expression it defines. For example: On the one hand, we have been trained to *express regret* for things that have happened, and – by inferential procedures whose nature is not well understood – to *ascribe regrets* to others. On the other hand, we have *not* been trained to say when a "desire to possess something" is being constrained, in the fashion Spinoza describes, by incompatible memories. Spinoza's definition, if we assume it to be correct, would tell us that on those more readily identifiable occasions on which the definiendum applies – that is, when someone regrets something – the definiens also applies. Think of Spinoza's definition of regret as an invitation to reexamine those occasions on which a person, oneself or another, has (as we assume) regretted something. The definition

is in part a hypothesis that on those occasions something like the following was going on:

- The "activation" of a memory of a goal that was once attainable – a memory of it *as attainable* – was sustaining (keeping alive) a desire to attain that goal. (For example, going for a sail on one's boat, or seeing one's spouse in the evening.)
- This in turn uncovered other memories, reminders that the goal is no longer attainable. (One has sold the boat, or it was stolen; one is divorced or widowed.)
- These memories led to the *recognition* that the goal is no longer attainable.
- This recognition was (as Spinoza adds) felt as a kind of "pain" (i.e., "feeling bad").

Of particular interest is the fact that Spinoza's definition portrays regret as a state possessing what one might call "causal depth"; that is, an instance of it occurs only when a state (or event) of one type S_1 *causes* a state (event) of another type S_2. The activation of a memory "keeps alive" – is a sustaining cause of – a certain desire; other memories cause us to recognize the unattainability of that desire; and that recognition causes pain. The hypothesis that this process goes on evokes a confirmatory "Aha!" in many readers, seemingly enabling them to "see," in their own episodes of regret and in those of others, a *dynamics* that had not been apparent before. Further, to say that human beings are susceptible to having regrets is (on Spinoza's definition) to say something fairly specific about "human nature." It is to say that in human beings certain states are *apt* to cause certain other states, for example, that the "activation" of a memory of a past goal *as attainable* is apt to reawaken a desire to attain that goal, and that given such a reawakening, a reminder that the goal is no longer attainable is apt to cause one to feel bad.

UNIVERSALITY AND FORMALITY

Whatever "insights" Spinoza's definitions give us into the etiology of particular emotions, they are strictly *universal* insights. To apply the insights to a particular individual A, we needn't know anything about A beyond the fact that A regrets something, or fears something, or whatever. We needn't know whether A is a man or a woman, an Earthling or an Andromedan, or even, perhaps, a distinctly nonhuman animal or robot to which we deem it fitting to ascribe regrets or fears.

It is further true that these definitional insights are, as I shall explain, *formal* in character. Notice, for example, the repeated reference to one and the same "thing" in Spinoza's definition of regret:

> Regret is the desire or appetite to possess *something*, kept alive by the remembrance of *the said thing*, and at the same time constrained by the remembrance of other things which exclude the existence of *it* [my ephasis].

What Spinoza is offering here is a form or template to be filled in for each instance of regret. This becomes clear if we replace the emphasized expressions '*something*,' '*the said thing*,' and '*it*' with a dummy variable '*x*':

> Regret is the desire or appetite to possess *x*, kept alive by the remembrance of *x*, and at the same time constrained by the remembrance of other things which exclude the existence of *x*.

Or in my reformulation:

> A desire to attain a particular *goal g* is being kept alive by the "activation" of a memory of *g as attainable*, which in turn awakens memories of something ("other things") that makes *g* now *un*attainable. (For example: *g* = going for a sail on one's boat, or *g* = seeing one's spouse in the evening.)

It is these "other things" that one mentions in specifying *what* one regrets – for example, that one has sold the boat, or that it was stolen; that one has gotten a divorce, or that one's spouse has died. Beginning with the "recognition" that Mary regrets they sold the boat, one may not know *why* she regrets this, that is, just what the relevant goal *g* is – she might regret it, after all, not because she'd like to sail it but because she'd like to pass it on to her children – but one can safely assume this much: It is some goal or goals that are (in Mary's view) *unattainable because they sold the boat.* This, of course, can be generalized in a formal way: If one regrets *y*, one does so because one has recurrent memories of some goal or goals that are (in one's view) unattainable because of *y*. Such generalizations open the way to what I shall call "formal insights" into the causes and effects of the various emotions.

AN APPLICATION OF FORMAL INSIGHT

Formal insight is exploited extensively in the following dialogue. This is a record of an actual interview in which a person (whom I shall call the "client") talks about one of his fears. Despite the

seemingly insightful conclusions at the end, it is important to notice that the interviewer exploits no prior knowledge of the client and indeed makes inferences solely on the basis of the client's answers. (The interviewer's comments and questions are numbered for reference in comments on the interview.)

1. What are you afraid of?
   Subways.
2. You have a general fear of subways. What is it about subways that you are particularly afraid of?
   I'm afraid of being mugged.
3. Afraid that you will be mugged?
   Yes.
4. What makes you think you might be?
   It's happened to people I know.
5. So that accounts for your fear. It would be unfortunate if you were to be mugged – and you can't be sure it won't happen, because, as you say, it's happened to people you know. Let's analyze your situation a little further. Is there anything you can do to make it *unlikely* that you'll be mugged? I mean within reason, of course.
   Yes.
6. What can you do?
   Avoid subways.
7. So you're afraid, really, of what would happen *if you didn't* avoid subways. That gives you a desire or tendency to avoid them. Sorry to ask you to spell this out, but it would help if you told me – just what is it about being mugged that you fear the most?
   I might be killed.
8. But isn't there something you can do – within reason – so that even if you were mugged, at least you wouldn't be killed?
   Yes.
9. What can you do?
   Carry a gun.
10. I see. So your fear of being mugged gives you some motivation to carry a gun. That would make you feel less vulnerable, wouldn't it? It would reduce your fear, make you less afraid of being mugged. That way, you wouldn't have to avoid subways. . . . By the way, this is not a recommendation. I'm much too stupid to tell you what to do. I don't even know what subways are. Or what *you* are, for that matter. After all, I'm just a computer – whatever that is.

COMMENTS ON THE INTERVIEW

As we learn from the final self-reflection, the interviewer is a computer, or more accurately a computer program. It was written to

demonstrate and test the analysis of fear offered in Chapter 4.[5] The above is a printout of a typical "interview." The interviewer is indeed rather "stupid" – by design, as well as by the severe limitations of my ability to program an "artificially intelligent" interviewer. As already stated, the interviewer exploits no prior knowledge of the client: For all it knows, the client might be a woman, a man, even an android or an Andromedan – so long as androids and Andromedans say they have fears. Its inferences are based solely on certain formal features of the client's answers – such as the fact that the answer to (1) is a single word ending with a single 's' (and is therefore almost certainly a plural noun), and that the word after 'afraid of' ends with an 'ing' not preceded by 'th,' and so on (and is therefore almost certainly a gerund formed from a verb).

The program is constitutionally unable to offer anything more than what I have called "formal insight" – insights into "what is going on" in the client that ought to be appropriate *no matter how the client fills in the blanks*. Yet I am inclined to think that, so far as it goes, my program discusses fears more insightfully than most people do.

It will be useful to examine the program's responses in some detail.

The entire point of (2) and (3) is to channel the fear "content" into sentential (or propositional) form. That is, they maneuver the client into describing his fear in terms of a particular sentence – specifically, the sentence that completes the form 'I am afraid that _____.' I shall explain below why it is crucially important to have a description that associates it with a particular sentence.

The client's assertion that he is afraid of *subways* tells us little more than that the blank in the form 'I am afraid that _____' is to be filled in with a sentence that is, explicitly or implicitly, *about* subways. (In the example, the relevant sentence, 'I will be mugged,'

5 Ager (1984) comments on the program: "Many of the precise arguments of the modern analytic tradition can find alternative expression in the idiom of algorithms and flowcharts. [Gordon 1980] describes the relationship between the content of an emotion and certain belief types in abstract quasi-algorithmic terms. His program is driven by that structure, leaving content to be filled in interactively. The program then is a testbed of the plausibility of the thesis when applied to various content and, not incidentally, a mechanism for exploring possible counterexamples." The program, entitled EMOTIONS, is written in BASIC. It borrows a trick or two from Joseph Weizenbaum's ELIZA.